Estudio de Análisis Exploratorio de los Datos (EDA): Proyecto Final de Máster en Big Data

Cátedra de Inteligencia Artificial y Big Data - Facultad de Ingeniería Universidad ORT Uruguay

### **Identificación del proyecto**

| Nombre del proyecto | Predicción de afluencia de pasajeros para mejorar la gestión del transporte urbano en Montevideo en días lluviosos. |
| --- | --- |
| Nombre, número de estudiante y correo de contacto de estudiantes | Christian Vera – 146159  chv.facu@gmail.com  Verónica González- 245393  verogonzalezaguirre@gmail.com |

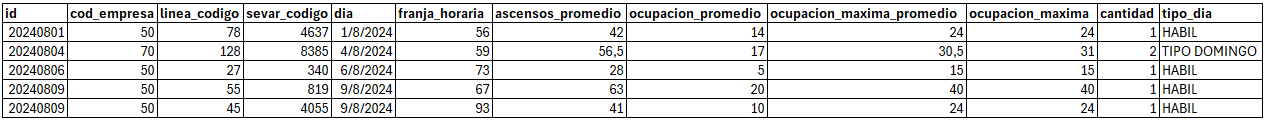
**Estudios realizados**

Se cuenta con los datos de las precipitaciones en la ciudad de Montevideo y la ocupación histórica de los ómnibus para cada línea. Estos datos son almacenados por la Intendencia de Montevideo, quienes han proporcionado los mismos. En una primera instancia hicimos pruebas con datos de un rango de fechas en ambos sets de datos que iban de 01/08/2024 al 31/10/2024. Pero notamos que necesitábamos más datos para poder tener un mejor panorama, por lo tanto se solicitó otro set de datos que va del 01/01/2024 al 01/01/2025.

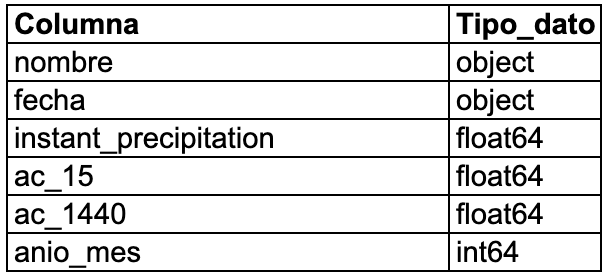
Información preliminar de dataset de ocupación histórica:

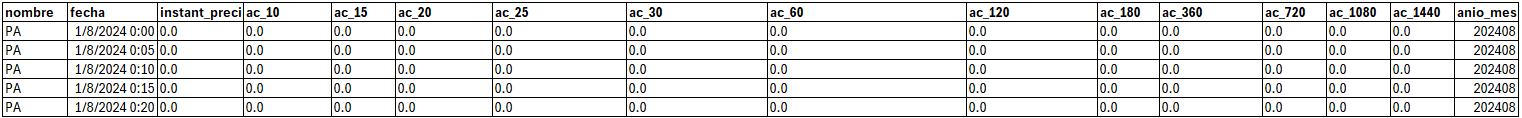
A table with text and numbers

Description automatically generated

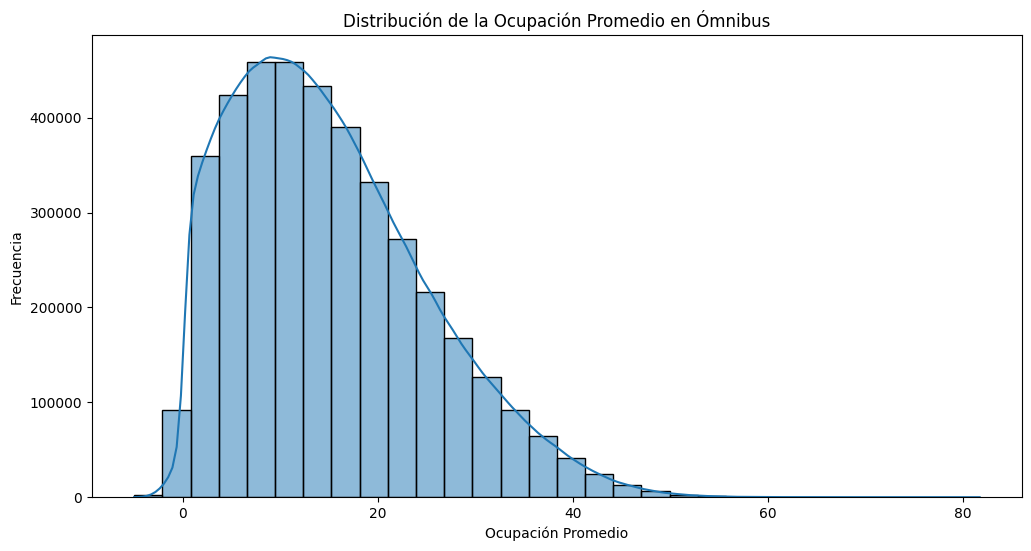


Información preliminar de dataset de precipitaciones:



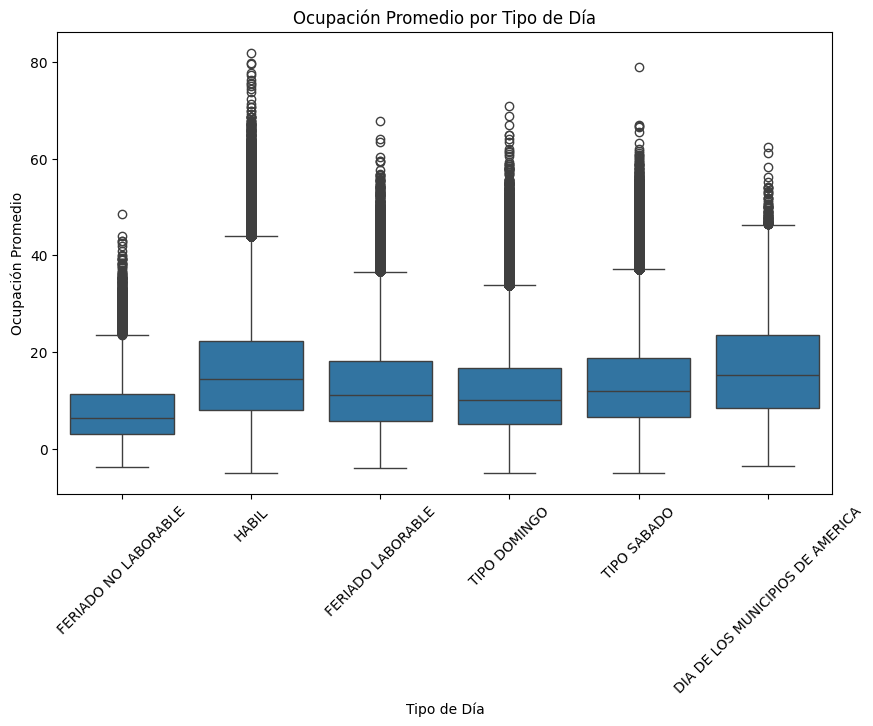


Se muestra la distribución de la ocupación promedio en los ómnibus:

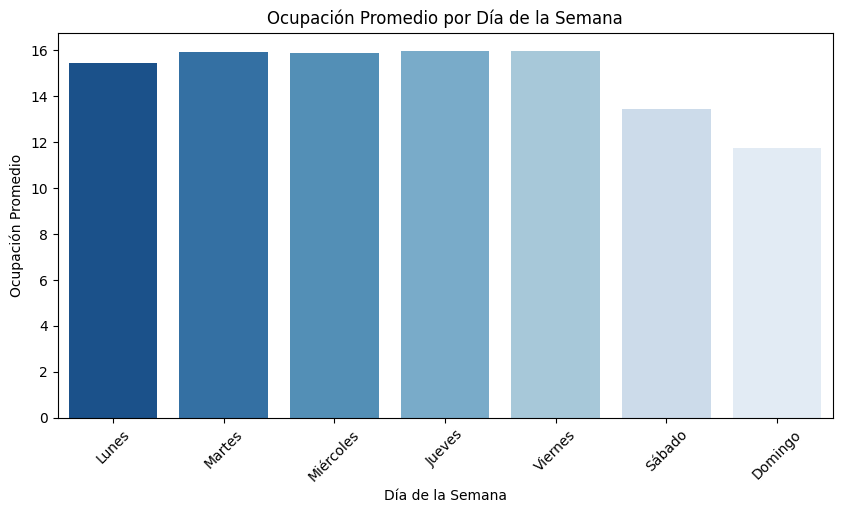


La distribución de la ocupación promedio en los ómnibus muestra una clara asimetría a la derecha, indicando que la mayoría de los viajes tienen una ocupación baja, mientras que existen algunos casos con valores significativamente más altos. La moda se encuentra en valores bajos, reflejando que la mayoría de los ómnibus viajan con ocupación reducida. Además, la cola larga hacia la derecha sugiere que en ciertos momentos o líneas específicas, la ocupación puede ser mucho mayor.

Ocupación promedio por tipo de día

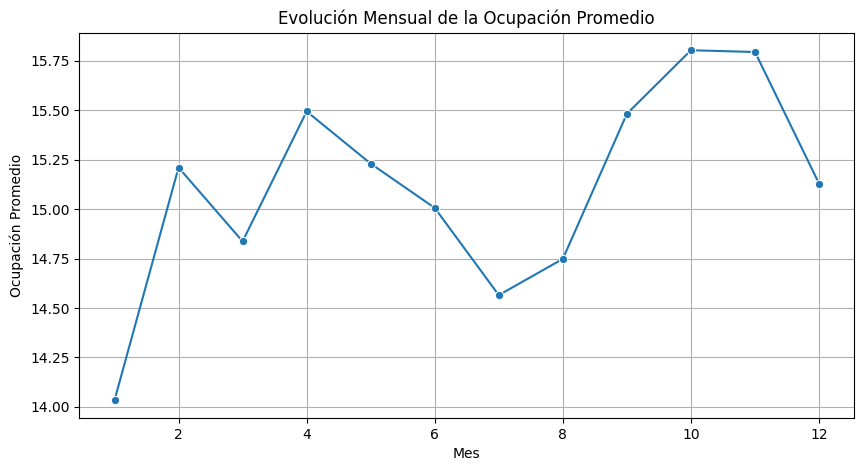


La ocupación promedio es más alta en días hábiles en comparación con los fines de semana y feriados. Se observa una mayor dispersión en los datos en días hábiles y feriados laborables, con valores atípicos que representan momentos de alta demanda. Los feriados no laborables tienen la menor ocupación promedio y menor dispersión, lo que indica menor uso del transporte en estos días, que también coincide con que estos días la frecuencia de ómnibus es mucho más reducida.

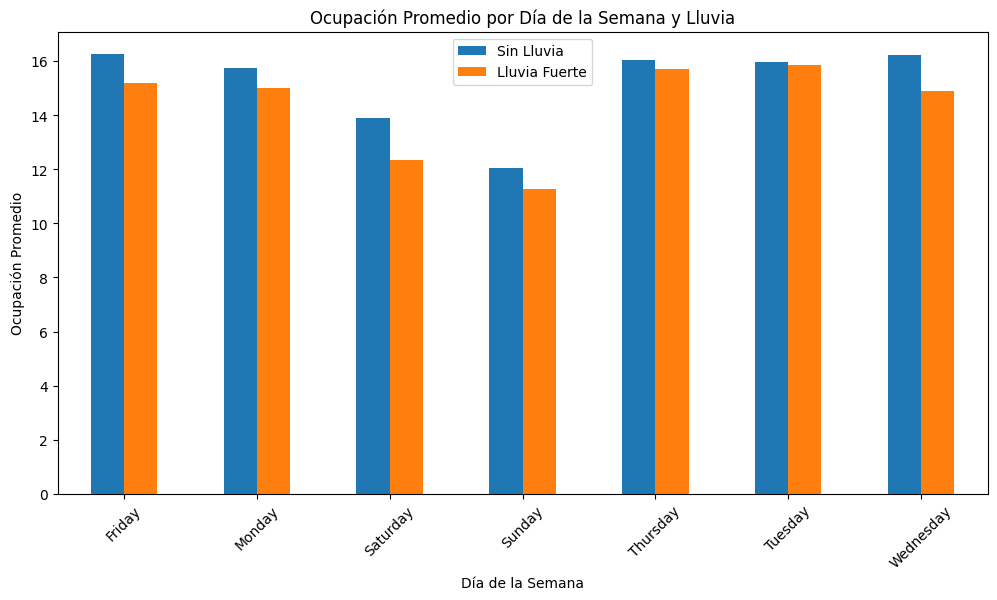
Ocupación promedio por día de la semana:  


Se puede visualizar que los días de semana (de Lunes a Viernes) la ocupación promedio es mayor a la de los Sábados y Domingos. El día que presenta la menor ocupación promedio es el Domingo.

Al analizar los mínimos de la ocupación promedio, ocupación máxima promedio y ocupación máxima se detectan valores negativos, se analizó la causa y se encontró dicho errores en los datos. Al consultar con los técnicos de la Intendencia de Montevideo sobre este problema, nos informan que los datos negativos no deberían de existir sino que deberían ser 0 dado que es un error por la forma de hacer el cálculo de ocupación. Recordar que los ómnibus en Montevideo solo cuenta con maquina de abordo al subir al mismo y la bajada se calcula, esto sería más exacto y no necesitaria de dicho cálculo si los ómnibus contarán con algún sistema al descender que descontará el pasajero que bajo como existe en otros países. Por esto se hizo la corrección de los datos pasando los datos de ocupación que eran negativos a 0



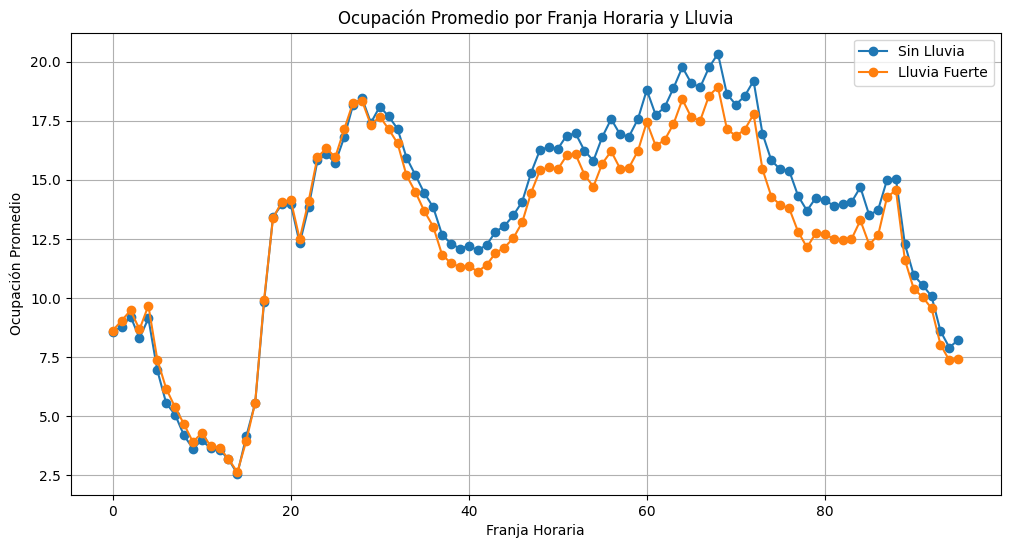
Al agregar más datos en una segunda instancia pudimos ver la evolución del promedio de ocupación en el año 2024 donde nota algunos picos pero un mayor crecimiento en los últimos meses del año.



Tratando de establecer una relación entre las ocupaciones promedio de los ómnibus y los días de lluvia, vemos que los días con lluvia presentan una menor ocupación promedio en todos los días de la semana, con una diferencia un poco más acentuada los fines de semana.

La correlación entre ocupación promedio y precipitación total es baja ( -0.05), lo que refuerza la idea de que la cantidad de lluvia no tiene un impacto fuerte directo sobre la ocupación en términos generales.

~~En días hábiles, la ocupación no parece verse tan afectada, lo que podría indicar que los desplazamientos por trabajo o estudio siguen siendo necesarios incluso en días lluviosos.~~



En esta gráfica se puede visualizar la ocupación con lluvia y sin lluvia a lo largo del día. Vale aclarar que la franja de tiempo se calcula como el número que aparece dividido cuatro porque está cada 15 minutos. ej 40/4= 10 osea que la franja 40 son las 10:00 am. También observamos que en las horas picos que es donde hay más frecuencia de ómnibus es donde más se separa una línea de la otra aunque sigue siendo muy poco la diferencia de ocupación con y sin lluvia.

**Conclusiones generales**

En una primera instancia se habían tomado 3 meses de datos, pero nos dimos cuenta que nos faltaba tener más datos, dado que no se podía tener una situación clara del comportamiento general de la ocupación, por este motivo solicitamos un año de datos. Se hicieron las pruebas con los datos de todo el 2024, y si bien no varió demasiado en comparación con lo que se tenía anteriormente, algunas cosas se pudieron ver más claras y no tan sesgadas. Se sigue observando que cuando llueve la ocupación disminuye en forma general aunque no de forma significativa. En los horarios que más se nota esta diferencia es en la hora “picos”. Esto último se pudo observar mejor al aumentar los datos.

Conclusión general: completaría el paso de extraer más datos, entiendo pertinente al menos un año, porque si nuestro enfoque y predicción será por día que parecería ser lo más lógico, nos quedamos con poca info agregada usando solo tres meses, además esos meses pueden ser especialmente secos o al revés, al usar un año tenemos más posibilidades de tener una foto más completa de la situación en todas las estaciones. Entiendo pertinente ver por qué las ocupaciones tienen valores negativos, conceptualmente no lo entiendo.